

Collections Digitization Framework: A Service-oriented Approach to Digitization in Academic Libraries

Karim Tharani
Information Technology Librarian
University of Saskatchewan
karim.tharani@usask.ca

Abstract

With advances in information technologies, academic libraries are now in a position to offer collections digitization services beyond campus communities. By mobilizing digitization to off-campus communities, academic libraries can reposition themselves as responsive and relevant in the face of a changing digital services landscape. This will also help academic libraries create unique opportunities to access and share hidden knowledge embedded within local and remote communities with rich intellectual traditions.

This article proposes a service-oriented framework for academic libraries to reimagine and mobilize collections digitization as part of broader library services. The proposed Collections Digitization Framework is based on the Service Framework for Digital Libraries developed by the Digital Library Federation, whereby digitization activities have been formalized into discrete processes and functions. The issues and challenges that academic libraries may face in mobilizing digitization services are also discussed in the context of a collaborative community digitization initiative undertaken by two Canadian academic libraries.

Keywords

community digitization; digitization activities; collaborative digitization; academic libraries; digitization services; Digital Library Federation; Service Framework for Digital Libraries; digital libraries

★ ★ ★

Academic libraries across North America continue to experience an increase in demand for digitization. Digitization activities in academic libraries, however, are usually managed as part of discrete projects rather than general library services (Mugridge 12). Services and projects play equally important and complementary roles in the successful functioning of academic libraries. A service is typically part of a library's core operations that has well-defined inputs and outputs. It is perceived as strategically valuable and is therefore offered to patrons consistently with costs and risks spread across the organization's operational budget. A project, on the other hand, is meant to be temporary in nature and has well-constrained timelines, resources, and associated risks to achieve desired outcomes. Being limited in time and resources, projects are often

also used "as the mechanism by which a service is deployed, retired, modified, or updated" (Scarborough).

Over time, libraries, archives, and museums have acquired extensive expertise and experience in managing digitization projects. In 2009 the Audio-Visual Working Groups produced the *Digitization Activities – Project Planning and Management Outline*¹ in which "the general sequencing of the work processes were identified using project management outlines from a number of organizations that had significant experience with the digitization of cultural materials." On the services side, the Digital Library Federation (DLF) has also developed the *Service Framework for Digital Libraries*² to view and transform disparate activities within libraries into modular services that support the overall mission.

With advances in information technologies, academic libraries are now in a position to reimagine digitization as part of library services. The structure provided by the DLF Service Framework is a useful resource to recast project-based digitization activities as library services. Thus with an overlay of the DLF Services Framework applied to the prevalent digitization activities, this paper proposes the Collections Digitization Framework for academic libraries to reimagine and mobilize collections digitization services.

The Collections Digitization Framework

Digitization is an overloaded term and is used in multiple contexts in academic libraries. For the purpose of the Digital Collections Framework, digitization refers to the reformatting of physical or analogue materials to create digital surrogates in order to provide access. The DLF Service Framework utilizes concepts such as business requirements, business processes, and business functions to map library activities to library services (see Figure 1). The DLF framework is also supported by concepts such as *business entities* and *abstract services*.

¹ For more information on the guideline, see "Guidelines: Digitization Activities - Project Planning." Still Image and Audio-Visual Working Groups. Federal Agencies Digitization Guidelines Initiative, 2009. Web. 26 Aug. 2012. <<http://www.digitizationguidelines.gov/guidelines/digitize-planning.html>>.

² For more information on DLF Service Framework, see Dempsey, Lorcan and Brian Dempsey. *A Progress Report for the DLF Steering Committee*. Digital Library Federation, 2005. Web. 26 Aug. 2012. <<http://old.diglib.org/architectures/serviceframe/dlfserviceframe1.htm>>.

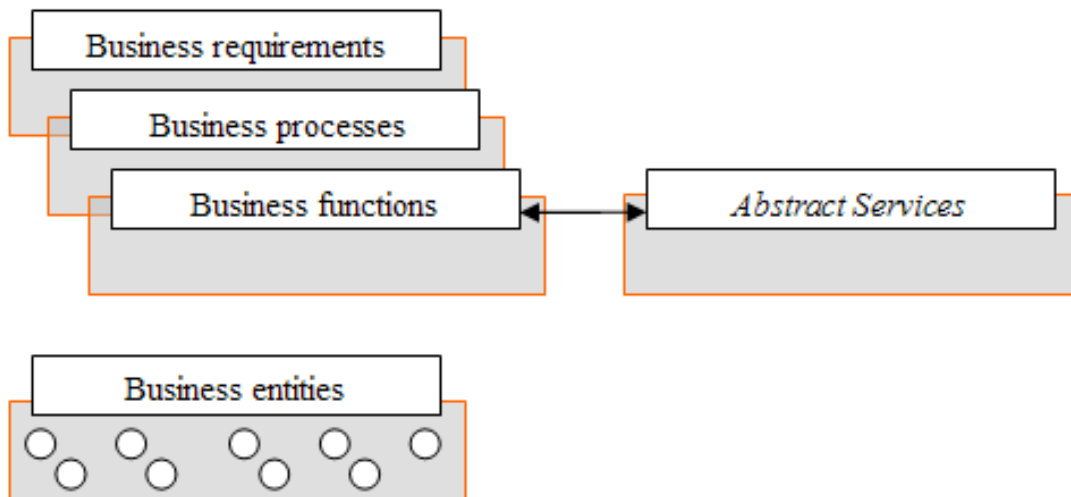


Figure 1 – The Service Framework for Digital Libraries (Lavoie, Henry, and Dempsey, 2006).

Business Requirement

At the core of the DLF Service Framework is the concept of the business requirement, which is defined as an "identifiable segment of an organization's overall mission" (Lavoie, Henry, and Dempsey). In the context of collections digitization, the business requirement is aptly articulated in the Digitization Activities – Project Planning and Management Outline as the:

complete process that broadly includes: selection, assessment, prioritization, project management and tracking, preparation of originals for digitization, metadata collection and creation, digitizing, quality management, data collection and management, submission of digital resources to delivery systems and into a repository environment, and assessment and evaluation of the digitization effort (4).

Business Processes and Functions

According to the DLF Service Framework, once the business requirement has been identified, it is broken into high-level business processes. In the proposed Digital Collections Framework, the collections digitization requirement has been further broken down into the following high-level business processes: Decide, Deploy, Describe, Digitize, Deposit, Display, and Direct (see Figure 2). Although the processes have been identified as discrete and sequential, in practice some of these activities may overlap or occur concurrently.

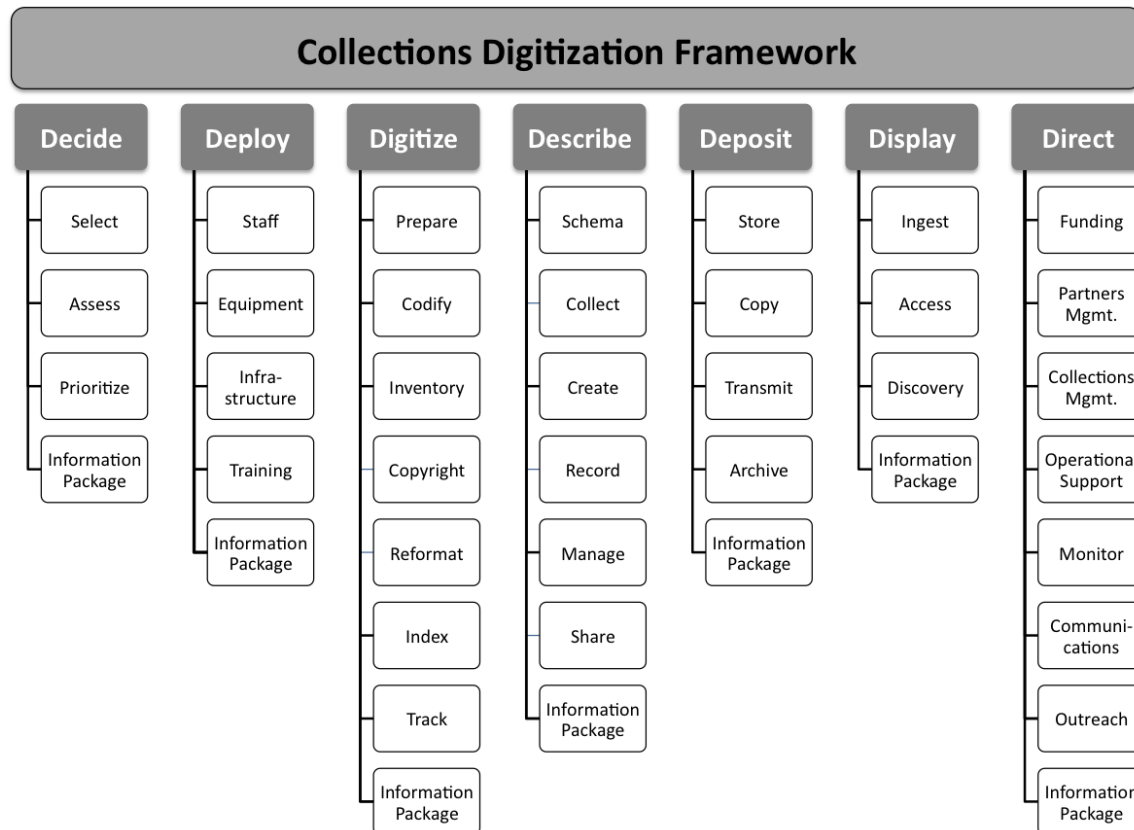


Figure 2 – The proposed Collections Digitization Framework

Each business process, in turn, is decomposed into discrete business functions. The business functions in the Digital Collections Framework, as described in Table 1, are designed to accommodate both single and multi-site digitization activities. With this accommodation, the proposed framework subscribes to the broader goal of providing a foundation for libraries to undertake large-scale collaborative activities and to mobilize their collective resources to respond promptly and efficiently to changing needs as envisioned in the DLF Service Framework (Dempsey and Dempsey).

Table 1 – Business Processes and Functions – Collections Digitization Framework

1. Decide: Qualify collections and items for digitization to support the library's mission.	
Select	Select candidate collections and items for digitization based on agreed upon criteria to support library goals.
Assess	Evaluate items based on their physical condition, copyright restrictions, provenance, availability of existing records, etc.
Prioritize	Rank selected items for digitization based on agreed upon criteria.

Information Package	This information package is used for capturing pertinent information about selected items from owners, producers and selectors, which is passed on to subsequent business processes and entities.
---------------------	---

2. Deploy: Identify and mobilize required resources to support digitization activities.

Staff	Identify and deploy the necessary human resources for digitization.
Equipment	Secure and deploy required equipment (e.g., workstations, scanners, etc.) to support digitization.
Infrastructure	Deploy the necessary infrastructure components such as network and FTP (file transfer protocol) connectivity, user accounts, file server, etc.
Training	Determine and address the training needs of staff engaged in digitization activities.
Information Package	This information package may be used to capture and communicate pertinent information about equipment, infrastructure, staff skills, as well as training.

3. Digitize: Generate digital surrogates for physical items selected for digitization.

Prepare	Organize items for digitization by determining appropriate reformatting techniques, formats, file naming conventions, etc.
Codify	Perform any linguistic or other processing that may be required e.g., transliteration or translation of foreign language items.
Inventory	Create and manage the inventory of items to be digitized by assigning them unique identifiers.
Copyright	Determine and comply with applicable copyright and intellectual property rights of the physical items, and formulate appropriate access and use restrictions pertaining to digital surrogates.
Reformat	Reformat physical items to create their digital surrogates. This function may also include post-reformatting corrective and quality control measures necessary to ensure desired quality of digital surrogates.

Index	Take necessary measures such as optical character recognition (OCR) to support subsequent use of digital surrogates.
Track	Track reformatting status of items based on agreed upon methods.
Information Package	This information package function may be used for capturing information about digital surrogates for subsequent processing.

4. Describe: The creation, collection, dissemination, and management of metadata to describe digitized items.

Schema	Identify appropriate metadata standards for the items to be digitized.
Collect	Gather existing information for the selected items.
Create	Create descriptive and other necessary metadata for the selected items.
Record	Electronically capture required and necessary metadata during digitization activities for management and reuse.
Manage	Manage existing and new metadata produced or consumed during digitization. This function is also responsible for linking metadata to digital surrogates.
Share	Prepare and share metadata and re-purpose metadata to other desired metadata schemas.
Information Package	This information package facilitates exchange of metadata between relevant business entities and processes.

5. Deposit: Facilitate submission of digital surrogates to appropriate local, archival, or access servers and repositories.

Store	Store digital files in accordance with agreed upon guidelines and standards.
Copy	Create additional copies of digital files as needed for sharing and safekeeping.
Transmit	Ensure that desired digital files are successfully transmitted, whether electronically or otherwise, to desired target locations.

Archive	Safeguard digital files based on agreed upon policies and procedures.
Information Package	This information package is utilized for submission of digital files and metadata to appropriate target servers and systems for storage.

6. Display: Provide support for presenting, searching, and discovery of digital items through appropriate servers and systems.

Ingest	Prepare and transfer digital files to relevant target servers and systems for the purpose of facilitating access.
Access	Verify availability and accessibility of necessary digital files based on agreed upon access criteria.
Discovery	Make relevant metadata available to facilitate search and retrieval of digital items.
Information Package	This information package is used for sharing digital files and associated metadata to enable access, display, and reuse of digital items on other target servers and systems as necessary.

7. Direct: Planning, administration, and monitoring of the digitization activities.

Funding	Secure funding for digitization initiatives and allocate resources and budgets.
Partners Management	Identify, qualify, and manage partners for planning and implementation of digitization initiatives.
Collections Management	Manage collections during and beyond digitization activities including policies, preservation, etc.
Operational Support	Provide operational support (people, process, and technology) to staff and stakeholders as necessary.
Monitor	Perform quality assurance and assessment of processes and functions to support digitization activities.
Communications	Provide timely and relevant information to stakeholders using appropriate means and messages. The responsibility of meeting any agreed upon reporting and status update requirements also

	falls under this function.
Outreach	Promote existing and planned digitization initiatives to scholars, students, exhibitors, government, private funding agencies, etc.
Information Package	This information package may be used to capture and communicate pertinent information about various administrative functions such as selection and funding criteria, guiding policies and procedures, evaluation and assessment checklists, performance and status reports, as well as outreach materials.

Business Entities

Business entities are information items, usually in the form of agreed upon content and format, deemed useful within the context of digitization activities. Examples of business entities may include information about providers, consumers, policies, metadata schemas, etc. In this regard, the proposed Digital Collections Framework follows the lead of the Open Archival Information Systems (OAIS) Reference Model³, which makes use of various *information packages* to facilitate flow and exchange of data among systems, stakeholders, and their environment at different stages of the digital preservation process. In the proposed framework, all processes have an *information package* function to capture, disseminate, and communicate pertinent information among business processes and entities. The actual content, structure, and frequency of a specific information package for a given initiative can be determined by stakeholders to address local needs and constraints.

Abstract Services

An abstract service, according to the DLF Service Framework, is a logical business functionality that could be automated and deployed on local infrastructure. Information packages are good candidates for abstract services as they can be developed and shared based on local requirements and infrastructure using the concepts of "service binding" and "deployed service" as specified in the DLF Service Framework.

Mobilizing Digitization Services

The University of Saskatchewan (U of S) Library funded a collaborative research initiative in order to identify and explore issues and considerations that academic libraries may face in mobilizing digitization to off-campus communities. For this research, the U of S Library selected the collection of the late Dr. Alwaiz Abualy (1919 – 2008), a prominent scholar of Islam and a theologian and preacher of considerable distinction among the Ismaili Muslims. The Abualy collection is especially rich in difficult

³ For more information on the OAIS Reference Model, see Lavoie, Brian F. OCLC Technology Watch Report. *The Open Archival Information System Reference Model: Introductory Guide*. Jan. 2004. Web. 30 Aug. 2012. <http://www.dpconline.org/component/docman/doc_download/91-introduction-to-oais-introduction-to-oais>.

to find South Asian Ismaili literature in Gujarati and Khojki scripts. This collection is also of particular interest to the Centre for the Comparative Study of Muslim Societies and Cultures at Simon Fraser University (SFU).

As part of this initiative, the U of S Library partnered with the SFU Library and the Abualy family in British Columbia to digitize and describe the Abualy Collection. Based on this tripartite partnership, the Abualy family provided access to the collection, and the two academic libraries collaborated to provide digitization services. In particular, the U of S Library agreed to provide the funding and operational administration, including hiring and training research assistants as well as deploying necessary digitization equipment. The SFU Library agreed to ingest and manage digitized items in its digital repository and to make the collection accessible on the Web. Due to its collaborative setup, this initiative can provide useful insights for exploring the challenges, issues, and decisions that academic libraries may face in transforming and mobilizing collections digitization as part of library services.

With more than two thousand items in the collection, the available funding was not sufficient to digitize the Abualy Collection in its entirety. This is not an uncommon situation faced by academic libraries engaged in digitization projects. In such situations, it is important that "libraries be clear about the purpose of a digital collection – whether it be for preservation, outreach, or curricular development; that libraries develop protocols for selection; and that libraries clarify the target audience for a digital collection" (Lopatin 277). For the Abualy Collection, there were many factors that were considered by the partners including uniqueness and fragility, logical grouping, research interests, as well as cultural sensitivities surrounding the items. In consideration of all these factors, the partners decided that a subset of rare community-generated literature in Gujarati was best suited for digitization.

Since parts of the collection were fragile and not entirely catalogued, the family felt it was important to have the digitization performed onsite at the Abualy residence in Burnaby, British Columbia. The family provided the necessary space to set up the digitization site. The digitization equipment (scanner, workstation, software, storage media, etc.) was subsequently deployed to the designated digitization site. As Middleton pointed out "some institutions may be too remote to take advantage of digitization centers, and others may hesitate to transport their delicate and valuable collections... [so] programs that want to encourage participation from small institutions may want to consider offering some flexibility in where items are digitized" (148). This is equally true for communities that may wish to share their hidden collections but either have concerns or lack the wherewithal to do so.

Upon general agreement of the materials to be digitized, the ultimate responsibility to select items for digitization was assigned to the Abualy family. The decision to let the Abualy family maintain control over the selection process was guided by the observations made on other similar community projects. For example, in the context of digitally preserving indigenous cultural knowledge, a United Nations report observed that when community partners are involved and empowered during the selection process, "they are able to choose and select the information, expressions and rituals

that they perceive as important, wish to preserve and pass on. They are also encouraged to recognise their own cultural heritage as being worthy of preserving and passing on to future generations" (Czermak, Delanghe, and Weng 6). Therefore it is important to involve community partners and observe sensitivity in such projects.

One of the issues encountered in the process of describing the Abualy Collection was the selection of an appropriate metadata standard. In a community project like this, academic libraries are well positioned to provide their expertise in matters related to metadata and digital content management. For example, based on the specific needs of a community or project, academic libraries can help stakeholders decide between "a simple metadata format with a minimum of fields and little or no substructure such as Dublin Core and a very detailed format with many data elements having various structural complexities such as MARC 21" (Guenther and McCallum 12).

Raising awareness of sound digital stewardship practices is another area where academic libraries can add value to community projects. Whether digitizing for access or preservation, all digitization initiatives must plan for ample storage and backup of digital files that are generated during the digitization process. In the context of the Abualy Collection, the digital files produced at the digitization lab were regularly copied over to portable digital media for redundancy. Also, copies of these digital files were shipped to the U of S Library on a regular basis. They were then logged and stored for safekeeping at the U of S Library.

Making library technology infrastructure available to community partners is also an area where academic libraries can contribute. For example, the SFU Library was able share its existing processes and systems for providing digital access to the Abualy Collection, which made the uploading, ingestion, and display of digital surrogates relatively easy and straightforward.

The need for continued administrative and management support was critical for the partners throughout the Abualy Collection digitization. As emphasized in the Digitization Activities – Project Planning and Management Outline, "[the] activities such as project management and tracking, quality management, process improvement, as well as metadata management/collection (of all types of metadata) are ongoing processes that will continue throughout the entire digitization project" (12).

Although the Abualy Collection digitization initiative was a unique and collaborative undertaking, the partners were cognizant of the potential operational challenges a multi-site initiative like this could encounter. This led to multiple planning meetings among the partners to negotiate and define individual responsibilities, project milestones, and communication points for the project to ensure timely and efficient execution. As Middleton explains:

... the most challenging aspects of collaborative digitization programs are social rather than technical. Even at the institutional level, digitization programs often require that staff from different departments work with one another much more closely than they are accustomed. This challenge is compounded when a digitization program includes multiple institutions of various types, sizes, and

locations. ... By respecting these differences and insuring that all participants benefit, digitization programs can develop collections that are more diverse, increase the efficiency of the digitization process, and create additional funding options (146).

Conclusion

The Collections Digitization Framework provides and instills a common understanding and vocabulary for digitization processes and functions. This is an important first step; however, it is equally important for academic libraries to transform and offer digitization as part of general library services. Collaborating to mobilize digitization services beyond campus communities can also help advance scholarship and learning by ensuring consistent discovery of hidden collections, such as the Abualy Collection. As pointed out by the Council on Library and Information Resources,

... the impact and influence of this program could be profound. On one level, it will make visible and accessible valuable collections that would otherwise remain out of reach to scholars and students. By revealing these materials, the program may instigate new collaborations among scholars and facilitate new approaches to research: new questions may be asked, and new kinds of queries can be made against the data (11).

Partnering to fulfill their respective mandates is not a new concept in academic libraries. Interlibrary Loan (ILL) programs and various other library consortia are successful exemplars of academic libraries joining forces for mutual benefit with creativity and efficiency. As noted in the Roundtable on Technology and Change in Academic Libraries, convened by the Association of College and Research Libraries,

[t]he challenge for libraries, their leadership and staff, is to recast their identities in relation to the changing modes of knowledge creation and dissemination, and in relation to the academic communities they serve. ... Libraries must descend from what many have regarded as an increasingly isolated perch of presumed privilege and enter the contentious race to advance in the market for information services—what one participant in our roundtable termed '*taking it to the streets*' (Wegner and Zemsky 5).

It is therefore critical for academic libraries to realize the strategic value of digitization activities in the context of their services and mission. Given the changing digital services landscape, repurposing existing digitization expertise and experience may be an opportune endeavour for academic libraries to undertake in order to remain responsive and relevant. The proposed Collections Digitization Framework is one attempt to reimagine and mobilize collections digitization as part of broader academic library services.

Acknowledgements

The author is grateful to Dr. Vicki Williamson, Dean of the University Saskatchewan Library, for providing the funding for this research. The author also gratefully acknowledges the support and contributions of the following partners and colleagues during the planning and digitization of the Abualy collection:

- Mohamed Alibhai – Abualy Family Representative, Burnaby B.C.
- Mark Jordan – Head of Library Systems, SFU Library
- Derryl MacLean, – Associate Professor, History, SFU.
- Noorali Meghani – Research Assistant, U of S Library
- Brian Owen – Associate University Librarian, SFU Library
- Ian Song – Digital Initiatives Librarian, SFU Library
- Eric Swanick – Gifts Librarian, SFU Library

Works Cited

- "Cataloging Hidden Special Collections and Archives: Building a New Research Environment." *Council on Library and Information Resources (CLIR)*. Web. 26 Aug. 2012.
- Dempsey, Lorcan and Brian Dempsey. *A Progress Report for the DLF Steering Committee*. Digital Library Federation, 2005. Web. 26 Aug. 2012.
- Czermak, Karin, Philippe Delanghe and Wei Weng. UNESCO. *Preserving intangible cultural heritage in Indonesia: A pilot project on oral tradition and language preservation*. Jakarta: UNESCO Publishing, 2003. Web. 22 Jun. 2012.
- Guenther, Rebecca, and Sally McCallum. "New metadata standards for digital resources: MODS and METS." *Bulletin of the American Society for Information Science and Technology*. 29.2 (2003): 12-15. Web. 27 Jun. 2012.
- "Guidelines: Digitization Activities - Project Planning." *Still Image and Audio-Visual Working Groups*. Federal Agencies Digitization Guidelines Initiative, 2009. Web. 26 Aug. 2012.
- Lavoie, Brian, Geneva Henry and Lorcan Dempsey. "A service framework for libraries." *D-Lib Magazine*. 12.7/8 (2006). Web. 27 Jun. 2012.
- Lopatin, Laurie. "Library digitization projects, issues and guidelines: A survey of the literature." *Library Hi Tech* 24.2 (2006): 273-289. *Emerald Group Publishing Limited*. Web. 27 Jun. 2012.
- Middleton, Ken. "Collaborative digitization programs: a multifaceted approach to sustainability." *Library Hi Tech* 23.2 (2005): 145-150. *Emerald Group Publishing Limited*. Web. 27 Jun. 2012.

Mugridge, Rebecca L. *SPEC Kit 294: Managing Digitization Activities*. Washington, DC: Association of Research Libraries. 2006. Web. 22 Jun. 2012.

Scarborough, Michael. "The difference between a project and a service." *The Project Management Hut*. 2012. Web. 30 Jul. 2012.

Wegner, Gregory and Robert Zemsky. *Changing roles of academic and research libraries*. Roundtable on Technology and Change in Academic Libraries, Association of College and Research Libraries. 2007. Web. 22 Jun. 2012.